Bayesian Inference for Correlations in the Presence of Measurement Error and Estimation Uncertainty

Dora Matzke[1], Alexander Ly[1], Ravi Selker[1], Wouter D. Weeda[2], Benjamin Scheibehenne[3], Michael D. Lee[4], and Eric-Jan Wagenmakers[1]

[1] University of Amsterdam
[2] Leiden University
[3] University of Geneva
[4] University of California, Irvine

Correspondence concerning this article should be addressed to:
Dora Matzke
University of Amsterdam, Department of Psychology
PO Box 15906
1001 NK, Amsterdam, The Netherlands
Phone: +31205258862
E-mail to d.matzke@uva.nl.

Abstract

Whenever parameter estimates are uncertain or observations are contaminated by measurement error, the Pearson correlation coefficient can severely underestimate the true strength of an association. Various approaches exist for inferring the correlation in the presence of estimation uncertainty and measurement error, but none are routinely applied in psychological research. Here we focus on a Bayesian hierarchical model proposed by Behseta, Berdyyeva, Olson, and Kass (2009) that allows researchers to infer the underlying correlation between error-contaminated observations. We show that this approach may be also applied to obtain the underlying correlation between uncertain parameter estimates as well as the correlation between uncertain parameter estimates and noisy observations. We illustrate the Bayesian modeling of correlations with two empirical data sets; in each data set, we first infer the posterior distribution of the underlying correlation and then compute Bayes factors to quantify the evidence that the data provide for the presence of an association.

Keywords: **Attenuation of the correlation, Bayesian inference, cumulative prospect theory, diffusion model, estimation uncertainty, measurement error**

Acknowledgment

Introduction

A growing number of researchers rely on formal mathematical models for analyzing data obtained from psychological experiments. These models come in many flavors, ranging from simple statistical distributions, such as the ex-Gaussian (Matzke & Wagenmakers, 2009), to sophisticated cognitive process models, such as the diffusion model (Ratcliff, 1978). Regardless of its degree of sophistication, the general goal of mathematical modeling is to capture regularities in the data using parameters that represent separate statistical components or distinct psychological variables.

Typically, investigators estimate model parameters for each participant separately, for instance with maximum likelihood (e.g., Myung, 2003) or Bayesian methods (e.g., M. D. Lee & Wagenmakers, 2013). Once the parameter estimates have been obtained, researchers are sometimes interested in assessing the association between parameters, or between parameters and other observed variables. For example, a researcher may fit the diffusion model to the individual data sets obtained from a lexical decision task and compute the correlation between parameter estimates that reflect participants' response caution and parameter estimates that quantify participants' rate of information accumulation (e.g., drift rate). Alternatively, a researcher may be interested in computing the correlation between estimates of drift rate and measurements of general intelligence.

The Pearson product-moment correlation coefficient is a popular measure of the linear relationship between two variables (for Bayesian solutions, see Ly, Verhagen, & Wagenmakers, 2016a, 2016b; Ly, Marsman, & Wagenmakers, 2015). Its popularity in psychology is illustrated by the fact that 42% of the 67 articles in the 2012 volume of the *Journal of Experimental Psychology: General* (JEP:G) report at least one Pearson correlation coefficient, with 152 correlations in total, and an average of more than 5 correlations per article. Despite its popularity, psychology as a field seems generally unaware that the correlation coefficient can severely underestimate the true strength of the association in the presence of measurement error or estimation uncertainty.

Measurement Error

It is generally recognized that most—if not all—psychological constructs are measured imperfectly. It is also well-known among statisticians that measurement error decreases the correlation coefficient between two sets of observations (e.g., Charles, 2005; Spearman, 1904). Although various approaches are available to obtain correlation coefficients that take into account measurement error, none are routinely applied in psychology. Indeed, out of the 28 JEP:G articles in the 2012 volume that reported one or more correlations, only one acknowledged the deleterious effects of measurement error and corrected the observed correlation.

Attempts to remedy the problem of the attenuation of the correlation date back to Spearman (1904), who proposed to correct the correlation coefficient using the reliability with which the observations were obtained. Spearman's attenuation formula is related to errors-in-variables models, which are extensions to standard regression models that aim to correct the bias in parameter estimates that results from measurement error (e.g., Buonaccorsi, 2010; Cheng & Van Ness, 1999; Fuller, 1987; for Bayesian solutions, see Congdon, 2006; Gilks, Richardson, & Spiegelhalter, 1996; Gustafson, 2004; Lunn, Jack-

son, Best, Thomas, & Spiegelhalter, 2012; Muff, Riebler, Held, Rue, & Saner, in press). If both the criterion and the response variables are assumed to be measured with noise, Spearman's method and the correction within standard linear regression models result in the same disattenuation (see Behseta et al., 2009, Appendix). Cole and Preacher (2014) discuss solutions to deal with measurement uncertainty in the context of path analysis.

Estimation Uncertainty

It is also generally recognized that if only limited data are available, model parameters are typically estimated with some degree of uncertainty. It is, however, rarely acknowledged that the correlation coefficient computed between two sets of uncertain parameter estimates can severely underestimate the true strength of the association between the parameters. This widespread neglect is puzzling because various approaches are available to obtain correlation coefficients that take into account the uncertainty of the parameter estimates. The Bayesian estimation of correlations between parameters in regression models has been illustrated by Gelman and Hill (2007, p. 279). Bayesian solutions for inferring the correlation in the context of cognitive models have been outlined by Klauer (2010), Matzke, Dolan, Batchelder, and Wagenmakers (2015), Rouder, Lu, Morey, Sun, and Speckman (2008), and Rouder et al. (2007). Monte Carlo methods that deal with the adverse consequences of estimation uncertainty in cognitive models have been described by Ratcliff and Strayer (2014).

Measurement Error and Estimation Uncertainty: A Bayesian Solution

Methods that remedy the attenuation of the correlation typically target uncertainty that results from either imperfect measurements or parameter estimation, but do not deal with both sources of uncertainty simultaneously. Here we attempt to address this limitation and describe a Bayesian method that allows researchers to infer the underlying correlation between (1) two sets of error-contaminated observations; (2) two sets of uncertain parameter estimates; and (3) a set of uncertain parameter estimates and a set of error-contaminated observations. Our approach is based on the Bayesian hierarchical method proposed by Behseta et al. (2009) that explicitly models measurement error, and so can be conceived as a principled method for "correcting" the correlation for the presence of noise in the data.

The outline of this article is as follows. First, we introduce the basic concepts of Bayesian parameter estimation and hypothesis testing. Second, we illustrate the consequences of measurement error for the computation of the correlation and show that a similar problem applies to correlations derived between uncertain parameter estimates. Third, we describe a Bayesian hierarchical approach for inferring the underlying correlation in the presence of measurement error (Behseta et al., 2009) and show that the same method can also aid the recovery of the underlying correlation between two sets of parameter estimates. Finally, we illustrate our approach with two empirical data sets: one focusing on the correlation between parameters of cumulative prospect theory (Tversky & Kahneman, 1992), and the other focusing on the correlation between general intelligence and the drift rate parameter of the diffusion model (Ratcliff, 1978).

Bayesian Inference: Parameter Estimation and Hypothesis Testing

Bayesian inference starts with a prior distribution for the parameter of interest. The prior distribution quantifies our belief about possible values of the parameter before any data have been observed. The prior distribution is then updated by incoming data to obtain the posterior distribution. The posterior distribution reflects everything we know about possible values of the parameter after the data have been observed. The central tendency of the posterior, such as the mean, is often used as a point estimate for the parameter. The dispersion of the posterior, such as the variance, quantifies the uncertainty of the estimate: the larger the dispersion, the greater the uncertainty in the estimated parameter. The present article focuses on the estimation of the posterior distribution of the correlation coefficient ρ.

The posterior distribution often cannot be derived analytically; rather it must be approximated using numerical sampling techniques, such as Markov chain Monte Carlo sampling (Gamerman & Lopes, 2006; Gilks et al., 1996). Parameter estimation for the Bayesian hierarchical model outlined here may proceed using standard Bayesian statistical software, such as WinBUGS (Bayesian inference Using Gibbs Sampling for Windows; Lunn et al., 2012; for an introduction for psychologists, see Kruschke, 2010, and M. D. Lee & Wagenmakers, 2013). The WinBUGS script is presented in the Appendix. Note that the WinBUGS script requires minimal, if any, modification to run under OpenBUGS (Lunn, Spiegelhalter, Thomas, & Best, 2009) or JAGS (Plummer, 2003, 2013). The Stan project (Stan Development Team, 2012) provides yet another alternative to obtain the posterior distributions of ρ.

Once the posterior distribution of ρ has been computed, we may quantify the evidence the data provide for the presence of an association using Bayesian hypothesis testing. In Bayesian hypothesis testing, the prior odds for two models $p(\mathcal{M}_0)/p(\mathcal{M}_1)$ are updated by incoming data to obtain the posterior odds $p(\mathcal{M}_0 \mid D)/p(\mathcal{M}_1 \mid D)$ of the models:

$$\underbrace{\frac{p(\mathcal{M}_0 \mid D)}{p(\mathcal{M}_1 \mid D)}}_{\text{Posterior odds}} = \underbrace{\frac{p(D \mid \mathcal{M}_0)}{p(D \mid \mathcal{M}_1)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(\mathcal{M}_0)}{p(\mathcal{M}_1)}}_{\text{Prior odds}} . \tag{1}$$

The change in odds brought about by the data is given by the ratio of the marginal likelihoods $p(D \mid \mathcal{M}_0)/P(D \mid \mathcal{M}_1)$, also known as the Bayes factor (BF; Jeffreys, 1961; Kass & Raftery, 1995). In the present situation, $\mathcal{M}_0$ instantiates the null hypothesis of the absence of an association ($\mathcal{H}_0$: $\rho = 0$) and $\mathcal{M}_1$ instantiates the alternative hypothesis of the presence of an association ($\mathcal{H}_1$: $\rho \neq 0$). We follow Jeffreys (1961) and assign the correlation ρ under $\mathcal{M}_1$ a default prior distribution that is uniform from -1 to 1. The Bayes factor quantifies the probability of the data under $\mathcal{H}_0$ relative to the probability of the data under $\mathcal{H}_1$. For instance, BF_{01} of 10 indicates that the data are 10 times more likely under $\mathcal{H}_0$ than under $\mathcal{H}_1$. Alternatively, BF_{01} of $\frac{1}{10}$ indicates that the data are 10 times more likely under $\mathcal{H}_1$ than under $\mathcal{H}_0$.

We compute Bayes factors for two-sided hypothesis tests using the Savage-Dickey density ratio (e.g., Dickey & Lientz, 1970; Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010; Wetzels, Grasman, & Wagenmakers, 2010). Applied to the present situation, BF_{01} is given by the ratio of the height of the posterior and the prior distribution of ρ under $\mathcal{H}_1$ at

$\rho = 0$:

$$BF_{01} = \frac{p(D \mid \mathcal{H}_0)}{p(D \mid \mathcal{H}_1)} = \frac{p(\rho = 0 \mid D, \mathcal{H}_1)}{p(\rho = 0 \mid \mathcal{H}_1)}. \tag{2}$$

The height of the uniform prior distribution under $\mathcal{M}_1$ equals $\frac{1}{2}$ on the entire range, in particular at $\rho = 0$. To compute the height of the posterior distribution, we first obtain samples from the posterior distribution of ρ. We then approximate the posterior distribution by fitting the posterior samples with a scaled beta distribution ranging from -1 to $+1$ with parameters α and β. Lastly, we evaluate the height of the scaled beta distribution at $\rho = 0$ using the obtained α and β parameters.

Once the two-sided Bayes factor has been obtained, we compute one-sided—order-restricted—Bayes factors as recommended by Morey and Wagenmakers (2014), namely by correcting the two-sided Bayes factor using the proportion of posterior samples that is consistent with the order-restriction. The R script (R Core Team, 2012) for the computation of the Bayes factor is available in the supplemental materials at `http://dora.erbe-matzke.com/publications.html`.

Attenuation of the Correlation

First we illustrate the deleterious consequences of measurement error for the computation of the correlation. We then show that a similar problem applies to correlations computed between two sets of parameter estimates that are obtained from scarce data.

The Consequences of Measurement Error

Let θ and β be unobserved random variables and let $\hat{\theta}$ and $\hat{\beta}$ be the observed, error-contaminated measurements:

$$\hat{\theta} = \theta + \epsilon_\theta \tag{3}$$

and

$$\hat{\beta} = \beta + \epsilon_\beta, \tag{4}$$

where ϵ_θ and ϵ_β are the measurement errors associated with θ and β, respectively. The measurement errors are assumed to be normally distributed with variance $\sigma^2_{\epsilon_\theta}$ and $\sigma^2_{\epsilon_\beta}$:

$$\epsilon_\theta \sim \text{Normal}(0, \sigma^2_{\epsilon_\theta}) \tag{5}$$

and

$$\epsilon_\beta \sim \text{Normal}(0, \sigma^2_{\epsilon_\beta}), \tag{6}$$

so that

$$\hat{\theta} \sim \text{Normal}(\theta, \sigma^2_{\epsilon_\theta}) \tag{7}$$

and

$$\hat{\beta} \sim \text{Normal}(\beta, \sigma^2_{\epsilon_\beta}). \tag{8}$$

The correlation between the *unobserved* variables θ and β with variances σ^2_θ and σ^2_β, respectively, is given by

$$\rho = \frac{\sigma_{\theta\beta}}{\sqrt{\sigma_\theta^2 \sigma_\beta^2}}. \tag{9}$$

Assuming that the measurement errors are uncorrelated with θ and β and with each other, the correlation between the *observed* variables $\hat{\theta}$ and $\hat{\beta}$ is given by

$$r = \frac{\sigma_{\theta\beta}}{\sqrt{(\sigma_\theta^2 + \sigma_{\epsilon_\theta}^2)(\sigma_\beta^2 + \sigma_{\epsilon_\beta}^2)}}. \tag{10}$$

It follows from Equation 9 and Equation 10 that the observed correlation r is always lower than the unobserved true correlation ρ due to the additional source of variability that results from measurement error (see also Behseta et al., 2009, Appendix).

The Consequences of Estimation Uncertainty

In many applications of mathematical modeling, researchers estimate model parameters for each participant separately with maximum likelihood or Bayesian methods. In situations with only a limited number of observations per participant with respect of the number of parameters, the model tends to overfit the data. The resulting parameter estimates are therefore often overdispersed relative to the true parameters that generated the data (Farrell & Ludwig, 2008; Rouder, Lu, Speckman, Sun, & Jiang, 2005; Rouder, Sun, Speckman, Lu, & Zhou, 2003). In fact, we can distinguish two sources of variability in parameter estimates: variability that reflects true individual differences in model parameters and variability that reflects the uncertainty of the estimates that results from estimating model parameters from limited information. The inflation of the variability of the estimates as a result of estimation uncertainty may in turn decrease the observed correlation; as shown in Equation 9 and Equation 10, all else being equal, the larger the variances, the smaller the correlation.

Example. Consider the following hypothetical scenario. A researcher is interested in the heritability of the latencies of speeded two-choice decisions. In particular, the researcher hypothesizes that fast response times (RTs) are positively correlated between pairs of homozygotic twins. The researcher tests 80 twin pairs in a simple letter discrimination task (e.g., Thapar, Ratcliff, & McKoon, 2003) with 100 trials. The researcher models the RT data of each participant with the ex-Gaussian distribution (Hohle, 1965; Matzke & Wagenmakers, 2009). The ex-Gaussian distribution is a popular RT distribution that results from the convolution of a normal and an exponential distribution; the ex-Gaussian μ parameter gives the mean of the normal component and is often used to quantify the latency of fast responses. The researcher obtains posterior distributions for the μ_{T1} (Twin 1) and μ_{T2} (Twin 2) parameter for each twin pair using Bayesian inference. The researcher then computes the mean of the posterior distribution of the parameters and uses these as point estimates. Lastly, the researcher obtains a Pearson correlation coefficient of $r = 0.75$ between the two sets of point estimates.

Suppose that—unbeknownst to the investigator—the true value of the μ_{T1} and μ_{T2} parameters that generated the data are known for each twin pair, and so is the true correlation between the parameters: $\rho = 0.86$. The gray points in the top rows of Figure 1 show

the point estimates $\hat{\mu}_{T1}$ and $\hat{\mu}_{T2}$ plotted against their true values. The gray vertical arrows indicate the range of the point estimates; the black horizontal arrows indicate the range of the true values. The arrows show that the point estimates are more variable than the true values. Note that this result also holds in the maximum likelihood framework. Most importantly, as shown in the bottom panel, the Pearson correlation coefficient computed between the point estimates $\hat{\mu}_{T1}$ and $\hat{\mu}_{T2}$ (i.e., dotted line at $\rho = 0.75$) underestimates the true correlation between μ_{T1} and μ_{T2} (i.e., dashed line at $\rho = 0.86$).

A Bayesian Solution to Remedy the Attenuation of the Correlation

In this section, we outline the Bayesian approach proposed by Behseta et al. (2009) that allows for the estimation of the underlying correlation between error-contaminated observations. We then show that the same Bayesian model can also aid the recovery of the underlying correlation between uncertain parameter estimates.

Inferring the Correlation in the Presence of Measurement Error

Behseta et al.'s (2009) method for inferring the underlying correlation in the presence of measurement error relies on Bayesian hierarchical modeling (e.g., Farrell & Ludwig, 2008; Gelman & Hill, 2007; M. D. Lee, 2011; Rouder et al., 2005). The graphical representation of the hierarchical model is shown in Figure 2. Observed variables are represented by shaded nodes and unobserved variables are represented by unshaded nodes. The graph structure indicates dependencies between the nodes (e.g., M. D. Lee & Wagenmakers, 2013)

Level I: Observed Data. As before, let θ and β represent the true values, $\hat{\theta}$ and $\hat{\beta}$ the corresponding observed values, and ϵ_θ and ϵ_β the normally distributed errors associated with θ and β, respectively. For each observation i, $i = 1, ..., N$, $\hat{\theta}_i$ and $\hat{\beta}_i$ are given by

$$\hat{\theta}_i \sim \text{Normal}(\theta_i, \sigma^2_{\epsilon_{\theta i}}) \tag{11}$$

and

$$\hat{\beta}_i \sim \text{Normal}(\beta_i, \sigma^2_{\epsilon_{\beta i}}). \tag{12}$$

The error variances $\sigma^2_{\epsilon_{\theta i}}$ and $\sigma^2_{\epsilon_{\beta i}}$ are assumed to be known a priori or are estimated from data. Note that the model does not assume homogenous error variances across the N observations. Instead, each observation i has its own error variance.

Level II: Inferred Variables. For each observation i, the inferred parameters $\boldsymbol{\eta}_i = (\theta_i, \beta_i)$ are assumed to follow a bivariate normal distribution, with means $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{\eta}_i \sim \text{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{13}$$

where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_\theta \\ \mu_\beta \end{pmatrix} \tag{14}$$

and

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_\theta^2 & \rho\sigma_\theta\sigma_\beta \\ \rho\sigma_\theta\sigma_\beta & \sigma_\beta^2 \end{pmatrix}. \tag{15}$$

Here ρ is the underlying correlation between θ and β that is not contaminated by measurement error.

The group-level means $\mathbf{\mu}$ and the three elements (i.e., σ_θ, σ_β, and ρ) of the variance-covariance matrix $\mathbf{\Sigma}$ are estimated from data and require prior distributions. We use the following prior set-up:

$$\mu_\theta \sim \mathrm{Normal}(0, \sigma_{\mu_\theta}^2), \tag{16}$$

$$\mu_\beta \sim \mathrm{Normal}(0, \sigma_{\mu_\beta}^2), \tag{17}$$

$$\sigma_\theta \sim \mathrm{Uniform}(0, b_{\sigma_\theta}), \tag{18}$$

$$\sigma_\beta \sim \mathrm{Uniform}(0, b_{\sigma_\beta}), \tag{19}$$

where $\sigma_{\mu_\theta}^2$, $\sigma_{\mu_\beta}^2$, b_{σ_θ} and b_{σ_β} are set to large values to create relatively uninformative prior distributions. We assume that all values of the correlation are equally likely a priori and specify the following prior for ρ (Jeffreys, 1961):

$$\rho \sim \mathrm{Uniform}(-1, 1). \tag{20}$$

A benefit of the hierarchical formulation is that the correlation ρ is automatically adjusted for the additional source of variability that results from the uncertainty of the observations. In particular, the observations $\hat\theta$ and $\hat\beta$—especially those that are measured with great uncertainty—are "shrunk" toward their corresponding group mean. The degree of shrinkage is determined by the relative uncertainty of the observations, so that the posterior means of the inferred parameters θ and β are given by a weighted combination of the observations and the weighted group means. The inferred parameters are therefore less dispersed than the observed data points, and the inferred correlation is typically higher than the correlation computed with the direct observations.

Inferring the Correlation in the Presence of Estimation Uncertainty

The Bayesian hierarchical model proposed by Behseta et al. (2009) may be also used to infer the correlation between uncertain parameter estimates. We focus on parameter estimates obtained with Bayesian inference because the resulting posterior distributions can be conveniently used to quantify estimation uncertainty.

In the first step, we infer the posterior distribution of the model parameters for each participant. We use the posterior mean as a point estimate for the parameters; we use the posterior variance as a measure of the uncertainty of the estimates. Moreover, we assume that the point estimates can be conceptualized as an additive combination of a true parameter value and a normally distributed displacement parameter (see Equations 11-12).[1]

---

[1]Naturally, this assumption is sensible only if the posterior distribution is approximately normally distributed.

In the second step, we use the graphical model in Figure 2 with two minor adjustments to infer the underlying correlation between the model parameters. First, we replace the observed measurements $\hat{\theta}_i$ and $\hat{\beta}_i$ with the point estimates (i.e., mean of the posterior distributions). Second, we substitute the measurement error variances $\sigma^2_{\epsilon_{\theta i}}$ and $\sigma^2_{\epsilon_{\beta i}}$ with the uncertainty of the point estimates (i.e., variance of the posterior distributions).

As a result of the hierarchical formulation, the point estimates are shrunk toward their corresponding group mean, where the degree of shrinkage is determined by the relative uncertainty of the estimates. The correlation ρ is therefore automatically adjusted for the additional source of variability that results from imperfect parameter estimation.

Example Continued. Aware of the disadvantages of parameter estimation with scarce data, the researcher sets out to use Bayesian hierarchical modeling to infer the underlying correlation between μ_{T1} and μ_{T2}. The researcher uses the mean of the posterior distribution of the individual parameters as point estimates ($\hat{\mu}_{T1}$ and $\hat{\mu}_{T2}$) and quantifies their uncertainty with the variance of the posteriors ($\sigma^2_{\epsilon_{\mu T1}}$ and $\sigma^2_{\epsilon_{\mu T2}}$). The researcher then feeds the posterior means and posterior variances into the hierarchical model in Figure 2 and adjusts the prior distribution of the group-level means and group-level standard deviations to match the measurement scale of the RT data.

The black points in the top rows of Figure 1 show the posterior means of the inferred μ_{T1} and μ_{T2} parameters plotted against their true values. The black vertical arrows indicate the range of the inferred parameters; the inferred parameters from the hierarchical analysis are less dispersed than the point estimates (i.e., gray points). Most importantly, as shown in the bottom panel, the posterior distribution of the correlation ρ from the hierarchical analysis more closely approximates the true correlation between μ_{T1} and μ_{T2} than the Pearson correlation coefficient r computed between the point estimates.

Empirical Examples

We now turn to two examples from the existing literature. In the first example, we demonstrate how our Bayesian approach can be used to infer the underlying correlation between two sets of parameter estimates. In particular, we assess the correlation between parameters of cumulative prospect theory obtained at two different time points. In the second example, we show how the Bayesian approach can be used to infer the underlying correlation between a set of parameter estimates and a set of error-contaminated measurements. In particular, we assess the correlation between the drift rate parameter of the diffusion model and general intelligence as measured by the Raven Progressive Matrices Test.

Both examples proceed as follows: We estimate the posterior distribution of the model parameters for each participant separately and use the mean of the posterior distributions as point estimates. To highlight the advantages of our approach, we then conduct two separate analyses. In the first analysis, we compute the posterior distribution of the *observed* correlation r by modeling the point estimates or the Raven scores directly using a bivariate normal distribution. In the second analysis, we take into account the uncertainty of the parameter estimates or the Raven scores and compute the posterior distribution of the *inferred underlying* correlation ρ. Finally, for both analyses, we compute the Bayes factor (BF_{0+}) to quantify the evidence the data provide for the presence of a positive association.

Example 1: Correlation between Parameters of Cumulative Prospect Theory

Our first example features a data set obtained from a decision-making experiment reported in Glöckner and Pachur (2012). The 64 participants were instructed to choose between monetary gambles in two experimental sessions. The two sessions were separated by one week and each consisted of 138 two-outcome gambles.

We model the observed choice data with cumulative prospect theory (CPT; Tversky & Kahneman, 1992) and focus on the δ parameter that governs how individual decision makers weight the probability information of the gambles: high values of δ indicate high degree of risk aversion. As the CPT parameters are assumed to be relatively stable across short periods of time, we examine the correlation between the δ parameters estimated from data collected during the two experimental sessions. We infer posterior distributions for the model parameters for each participant, separately for the two testing occasions, by adapting the model used by Nilsson, Rieskamp, and Wagenmakers (2011).[2] The prior distribution for δ covers a wide range of plausible values based on previous research.

Once the posterior distributions of the individual model parameters are obtained, we use the posterior means as point estimates for the δ parameters on the two testing occasions: $\hat{\delta}_1$ and $\hat{\delta}_2$. For the estimation of the inferred correlation ρ, we use the variance of the posterior distributions to quantify the uncertainty of $\hat{\delta}_1$ and $\hat{\delta}_2$: $\sigma^2_{\epsilon_{\delta 1}}$ and $\sigma^2_{\epsilon_{\delta 2}}$.

Results

The results are shown in Figure 3. Panel A shows a scatterplot between the point estimates $\hat{\delta}_1$ and $\hat{\delta}_2$. Panel B shows the same scatterplot but includes the corresponding estimation uncertainties $\sigma_{\epsilon_{\delta 1}}$ and $\sigma_{\epsilon_{\delta 2}}$. Panel C shows a scatterplot of the mean of the posterior distribution of the inferred $\boldsymbol{\eta} = (\delta_1, \delta_2)$ parameters. Panel D shows the posterior distribution of the observed (r) and the inferred (ρ) correlation.

As shown in Panel A, the Pearson correlation coefficient r between the point estimates is 0.62. If we take into account the uncertainty of the point estimates, the correlation increases substantially. Panel C shows that the posterior means of the inferred δ_1 and δ_2 parameters are shrunk toward their group mean and are associated very strongly. Panel D shows that the posterior distribution of the inferred correlation ρ is shifted to higher values relative to the posterior of the observed correlation r. In fact, after accounting for the uncertainty in $\hat{\delta}_1$ and $\hat{\delta}_2$, the mean of the posterior distribution of the correlation increases from 0.61 to 0.92.

One-sided Bayes factors indicate decisive evidence (Jeffreys, 1961) for the presence of a positive association for the observed r as well as the inferred ρ correlation; in both analyses, the data are more than $4,000,000$ times more likely under $\mathcal{H}_1$ than under $\mathcal{H}_0$. This result is visually evident from the fact that both posterior distributions are located away from zero such that their height at $\rho = 0$ is all but negligible.

The dramatic increase in the correlation is not unusual. Figure 4 shows the results of a simulation study investigating the magnitude of the attenuation for different values of the

---

[2]The CPT account of performance in the Glöckner and Pachur (2012) data set is merely an illustration; we do not suggest that the CPT with the present parameter setting provides the best description of the data of the individual participants. Note also that Glöckner and Pachur reported the results from fitting a slightly different model than the one used in the present article.

true correlation in a parameter setting that resembles the one found in the present data. We conducted five sets of simulations, each with a different value of the true correlation: 0.92 (i.e., the posterior mean of the inferred correlation ρ in the present data), 0.21, 0, -0.21, and -0.92. For each set of simulations, we generated $1,000$ synthetic data sets with $N = 64$, using the estimation uncertainties $\sigma^2_{\epsilon_{\delta 1}}$ and $\sigma^2_{\epsilon_{\delta 2}}$ and the posterior mean of the group-level μ_{δ_1}, μ_{δ_2}, $\sigma^2_{\delta_1}$, and $\sigma^2_{\delta_2}$ parameters.[3] We then computed the Pearson correlation coefficient r^* in each synthetic data set. The gray violin plots (e.g., Hintze & Nelson, 1998) show the distribution of the $1,000$ r^* values for the five levels of the true correlation.

Two results are noteworthy. First, all else being equal, the larger the absolute value of the true correlation, the larger the attenuation. This relationship is also evident from Equation 9 and Equation 10. Secondly, considering the very high inferred correlation of $\rho = 0.92$ in the present data set (upper arrow), the observed Pearson correlation of $r = 0.62$ (lower arrow) seems reasonable. In particular, the observed Pearson r between the point estimates $\hat{\delta}_1$ and $\hat{\delta}_2$ is well within the 2.5^{th} and 97.5^{th} percentile of the Pearson correlations coefficients r^* predicted by the simulations.

In sum, taking into account the uncertainty of the point estimates resulted in a dramatic increase in the correlation between CPT parameters; the mean of the posterior distribution of the correlation increased from 0.61 to 0.92. The Bayes factor indicated decisive evidence for the presence of a positive association for the observed as well as the inferred correlation.

Example 2: Correlation between Drift Rate and General Intelligence

Our second example focuses on the correlation between the drift rate parameter of the diffusion model and general intelligence as measured by the total score of the 20-min version of the Raven Progressive Matrices Test (Hamel & Schmittmann, 2006; Raven, Raven, & Court, 1998) in a data set collected by Weeda and Verouden (unpublished data). The data set featured RT and accuracy data from 51 participants. The stimuli were borrowed from the π-paradigm (Vickers, Nettelbeck, & Willson, 1972; Jensen, 1998), a speeded two-choice RT task and consisted of a series of images, each with one horizontal and two vertical lines (i.e.; two legs) that together formed the letter π, with one of the vertical lines longer than the other. The task was to indicate by means of a button press whether the left or the right leg of the π was longer. Task difficulty was manipulated on three levels—easy, medium, and difficult—by varying the difference between the length of the two legs.

We model the RT and accuracy data with the diffusion model (Ratcliff, 1978; Wagenmakers, 2009). The diffusion model provides a theoretical account of performance in speeded two-choice tasks. The four key parameters of the diffusion model correspond to well-defined psychological processes (Ratcliff & McKoon, 2008; Voss, Rothermund, & Voss, 2004): response caution, bias, the time taken up by encoding and motor processes, and the rate of information accumulation, which is the drift rate parameter of interest.[4] As drift rate v is often associated with higher cognitive functions and reasoning (i.e., Ratcliff, Schmiedek, & McKoon, 2008; Ratcliff, Thapar, & McKoon, 2010; Schmiedek, Oberauer,

---

[3]Note that this procedure is not the same as the posterior predictive assessment of a model's absolute goodness-of-fit (e.g., Gelman & Hill, 2007; Gelman, Meng, & Stern, 1996).

[4]In addition to these key parameters, the diffusion model also features parameters that describe the trial-to-trial variability of the key parameters.

Wilhelm, Suss, & Wittmann, 2007; van Ravenzwaaij, Brown, & Wagenmakers, 2011), we examine the correlation between drift rate and general intelligence.

We infer posterior distributions of the four key diffusion model parameters for each participant using the diffusion model JAGS module (Wabersich & Vandekerckhove, 2014). The prior distributions of the model parameters are based on parameter values reported in Matzke and Wagenmakers (2009). As drift rate is known to decrease with increasing task difficulty (e.g., Ratcliff & McKoon, 2008), we implement the following order-restriction: $v_\text{difficult} < v_\text{medium} < v_\text{easy}$. For simplicity, we focus on the average of the drift rate parameters across the three task difficulty conditions ($\bar{v}$). The remaining parameters are constrained to be equal across the conditions, and we set $z = \frac{a}{2}$.[5]

Once the posterior distributions of the model parameters are obtained, we use the posterior means as point estimates for average drift rate: $\hat{\bar{v}}$.[6] For the estimation of the inferred correlation ρ, we use the variance of the posterior distributions to quantify the uncertainty of $\hat{\bar{v}}$: $\sigma^2_{\epsilon_{\bar{v}}}$. For the Raven total score g, we assume homogenous error variance across participants and—for illustrative purposes—investigate how the inferred correlation changes as a function of the amount of measurement noise assumed in the data. In particular, we examine three scenarios: we assumed that 5%, 25%, and 55% of the total variance in Raven scores is attributable to measurement error, corresponding to excellent, acceptable, and very poor reliability, respectively.[7]

Results. The results are shown in Figure 5. Panels A to C show scatterplots between the point estimates $\hat{\bar{v}}$ and the Raven total scores $\hat{g}$. The gray lines show the corresponding estimation uncertainties $\sigma_{\epsilon_{\bar{v}}}$ and the standard deviation of the measurement error σ_{ϵ_g} for the 5%, 25%, and 55% noise scenarios. Panels D to F show scatterplots between the mean of the posterior distribution of the inferred $\boldsymbol{\eta} = (\bar{v}, g)$ parameters for the three scenarios. Panels G shows the posterior distribution of the observed (r) and the inferred correlations (ρ) for the three scenarios.

As shown in panel A, the Pearson correlation coefficient between the point estimates is 0.14. Panels D to F show that the posterior means of the inferred $\bar{v}$ and g parameters are slightly shrunk toward their group mean. The degree of shrinkage is determined by the amount of measurement noise assumed for g: the larger the noise, the larger the shrinkage. Panel G shows that the mean of the posterior distribution of the inferred correlation ρ (i.e., black posteriors) is progressively shifted to higher values relative to the posterior of the observed correlation r (i.e., gray posterior). As expected, the magnitude of the shift increases as the amount of measurement noise increases. Note, however, that the shift is

---

[5]The diffusion model account of performance in the Weeda and Verouden data set (unpublished data) is merely an illustration; we do not suggest that the diffusion model with the present parameter constraints provides the best description of the data of the individual participants.

[6]Note that the scale of both drift rate and general intelligence is bounded: $\bar{v}$ can take on values between 0 and 5.86 (i.e., prior range) and the Raven total score can take on values between 0 and 36. The use of the bivariate normal group-level distribution shown in Figure 2 is therefore theoretically unjustified. As a solution, we may transform the drift rate parameters and the Raven scores to the real line using a probit transformation. Additional analyses not reported here confirmed that using the transformed $\bar{v}$ and g values yields results that are very similar to the ones obtained using the untransformed drift rates and Raven scores. For simplicity, here we report the results of modeling the untransformed parameters.

[7]Note that this is only an illustration; the reliability of the Raven is well documented and is considered adequate (internal consistency ~ 0.90 and test-retest reliability ~ 0.82).

modest even if we assume that the Raven total score is an extremely unreliable indicator of general intelligence. The mean of the posterior distribution of the observed correlation r equals 0.13; the mean of the posterior of the inferred correlation ρ equals 0.14 in the 5% noise, 0.16 in the 25% noise, and 0.21 in the 55% noise scenario. Note also that the posterior of ρ is quite spread out, a tendency that becomes more pronounced as the amount of measurement noise in g increases.

One-sided Bayes factors indicate evidence for the absence of an association between drift rate and general intelligence for the observed as well as the inferred correlation. The evidence, however, is "not worth more than a bare mention" (i.e., $1 < BF_{01} < 3$), following the categorization of evidential strength provided by Jeffreys (1961). The BF_{0+} for the observed correlation r equals 2.13. The BF_{0+} for the inferred correlation ρ decreases from 2.00 in the 5% noise, to 1.75 in the 25% noise, and to 1.32 in the 55% noise scenario. Even with extremely unreliable Raven scores, the data are thus still more likely to have occurred under $\mathcal{H}_0$ than under $\mathcal{H}_1$. Note however that BF_{0+} of 1.32—or even BF_{0+} of 2.13—constitutes almost perfectly ambiguous evidence, indicating that the data are not sufficiently diagnostic to discriminate between $\mathcal{H}_0$ and $\mathcal{H}_1$. Inspection of the posterior distribution of the correlations suggests a similar conclusion: both the observed and the inferred correlations are estimated quite imprecisely.

Figure 6 shows the results of a simulation study investigating the magnitude of the expected attenuation for different values of the true correlation in a parameter setting that resembles the one found in the present data. Throughout the simulations, we assumed that 55% of the total variance of the Raven scores is attributable to error variance. We conducted five sets of simulations, each with a different value of the true correlation: 0.92, 0.21 (i.e., the posterior mean of the inferred correlation ρ in the present data set), 0, -0.21, and -0.92. For each set of simulations, we generated $1,000$ synthetic data sets with $N = 51$, using the estimation uncertainties $\sigma^2_{\epsilon_{\bar{v}}}$ and the posterior means of the group-level $\mu_{\bar{v}}$, μ_g, $\sigma^2_{\bar{v}}$, and σ^2_g parameters. We then computed the Pearson correlation coefficient r^* in each synthetic data set. The gray violin plots show the distribution of the $1,000$ r^* values for the five levels of the true correlation.

As before, all else being equal, the larger the absolute value of the true correlation, the larger the attenuation. Moreover, considering the relatively low inferred correlation of $\rho = 0.21$ in the present data (upper arrow), the observed Pearson correlation of $r = 0.14$ (lower arrow) seems perfectly reasonable. In fact, the median of the Pearson correlation coefficients r^* predicted by the simulations very closely approximates the observed Pearson r between $\hat{\bar{v}}$ and $\hat{g}$.

In sum, taking into account the uncertainty of the variables resulted in negligible increase in the correlation between drift rate and general intelligence; even when assuming unrealistically unreliable Raven scores, the mean of the posterior distribution of the correlation increased only from 0.13 to 0.21. The Bayes factor indicated evidence for the absence of an association between drift rate and general intelligence for the observed as well as the inferred correlations, regardless of the magnitude of the error variance assumed for g. The evidence for $\mathcal{H}_0$ was, however, only anecdotal, a result that is attributable to the substantial uncertainty in the estimated correlations.

Discussion

In the psychological sciences, few researchers acknowledge that the Pearson correlation coefficient may underestimate the true strength of the association between two sets of observations as result of measurement error. Even fewer recognize that the attenuation also plagues correlation coefficients derived from uncertain parameter estimates. Although various approaches are available to infer the correlation in the presence of measurement error or estimation uncertainty, none of these methods deal with both sources of noise simultaneously. The goal of the present paper was to outline a method that allows researchers to infer the underlying correlation between (1) two sets of error-contaminated observations; (2) two sets of uncertain parameter estimates; and (3) a set of noisy parameter estimates and a set of error-contaminated observations.

We illustrated the use of the Bayesian approach with two empirical data sets. In the first example, we assessed the correlation between parameters of cumulative prospect theory and demonstrated that taking into account the uncertainty of the parameter estimates can result in a dramatic increase in the inferred correlation: the mean of the posterior distribution of the correlation increased from 0.61 to 0.92. In the second example, we assessed the correlation between general intelligence and the drift rate parameter of the diffusion model, where we examined three scenarios: we assumed that 5%, 25%, and 55% of the total variance in Raven scores is attributable to measurement error. The estimated correlation increased only marginally when the uncertainty of the variables was taken into account; even with extremely unreliable Raven scores, the posterior mean of the correlation increased from 0.13 to only 0.21. Importantly, although the posterior mean of the inferred correlation increased slightly as a function of the amount of measurement noise in g, so did the posterior standard deviation. This result is only logical; if we measure both variables with a great degree of uncertainty, we cannot be confident about the underlying value of the correlation either.

Our approach for inferring correlations between model parameters is a straightforward application of the hierarchical method proposed by Behseta et al. (2009) for modeling measurement error. The present framework can be in fact viewed as a simple Bayesian structural equation model with two latent variables, each with a single indicator (e.g., S.-Y. Lee, 2007; Song & Lee, 2012). The original formulation of the model relies on a slightly different prior set-up than the one used in this article. In particular, Behseta et al. used an Inverse-Wishart distribution to model the variance-covariance matrix $\mathbf{\Sigma}$ of the latent variables, whereas we chose to model the individual components in $\mathbf{\Sigma}$ rather than $\mathbf{\Sigma}$ itself. We feel that the latter specification is more intuitive and flexible. It allows researchers to adapt the range of the prior for the group-level σ_θ and σ_β to the measurement scale of their variables and to express prior knowledge about the likely values of the inferred correlation in a straightforward manner.

Of course, we can take into account measurement error only when the error variance of the observations is known or when it can be estimated from data. Our investigation of the extent of the disattenuation as a function of the amount of measurement noise in the Raven scores served merely as an illustration. In real-life applications, the magnitude of the error variance should not be cherry-picked to obtain the desired (higher) correlation; rather it should be estimated from data. Note that the questionable practice of using

arbitrary reliability coefficients is one of the major objections to Spearman's (1904) original attenuation formula (e.g., Muchinsky, 1996). Luckily, our literature review suggested that the uncertainty of the observations may have been estimated for approximately 25% of the reported correlations. This is possible, for example, when each observation was derived as the average of multiple trials in a repeated measures design.

Similarly, we can infer the underlying correlation between two sets of parameter estimates only if the estimation uncertainty is known. Throughout this article we used Bayesian inference because the resulting posterior distributions can be automatically used to quantify the uncertainty of the parameter estimates. In particular, we treated the mean of the posterior distribution of the CPT and diffusion model parameters as point estimates and used the variance of the posterior distributions as a measure of the participant-specific estimation uncertainty.

With scarce data, however, the posterior distribution of the individual parameters can be sensitive to the choice of the prior. As a result, estimation uncertainty—expressed in terms of the variance of the posterior—can depend on the prior setting used to derive the individual estimates. It follows that the inferred correlation can only be interpreted given the particular prior setting used to obtain the individual parameter estimates. The present two-step framework should therefore be considered as a post-hoc method for inferring correlations: in the first step we obtain posterior distributions for the model parameters under a particular prior setting; in the second step we infer the correlation between the estimates given the estimation uncertainties obtained under the prior setting used in the first step. In the context of theory-laden models, such as the CPT and the diffusion model, we do not consider prior sensitivity as a shortcoming of the two-step approach; we consider the prior as part of the model which should come from theory just as the likelihood does (M. D. Lee & Vanpaemel, 2016; Vanpaemel & Lee, 2012).

Within the Bayesian framework, we are not limited to the two-step procedure illustrated here; Bayesian hierarchical modeling allows for the *simultaneous* estimation of the individual parameters and their correlations. In simultaneous hierarchical modeling, the individual parameters are assumed to follow a bivariate (or multivariate) normal distribution that is described by the group-level means and a variance-covariance matrix. The group-level and the participant-level parameters are estimated simultaneously from the data, where the group-level bivariate normal distribution acts as a prior to adjust extreme individual estimates to more moderate values. As a result of shrinkage of the individual parameters, the correlation automatically takes into account the uncertainty of the participant-level estimates. The Bayesian estimation of the covariance structure of model parameters is illustrated, for example, by Gelman and Hill (2007), Klauer (2010), Matzke et al. (2015), Rouder et al. (2008), and Rouder et al. (2007).

The two-step procedure and the simultaneous hierarchical framework assume different generative models for the data and as such constitute different approaches to modeling correlations. The two-step procedure assumes that the individual model parameters that generated the data are uncorrelated a priori. The simultaneous framework assumes a priori correlated parameters and automatically uses this prior information during the estimation of the individual parameters. We see merits in both approaches. The two-step approach is easy-to-use and can be applied when only point estimates—from Bayesian or non-Bayesian procedures—and the corresponding estimation uncertainties are available. Moreover, the

two-step approach can be useful in exploratory analyses of the correlation between model parameters and external observations. The simultaneous modeling framework has the potential to correct for bias in the individual parameter estimates that may result from fitting a simpler but unrealistic model with a priori uncorrelated parameters to scarce data. However, in the context of theory-laden models with meaningful priors, one should consider whether the prior set-up for the multivariate normal distribution used in the simultaneous framework adequately approximates the desired prior set-up for the cognitive model in question. Note that the two-step and simultaneous modeling frameworks yield similar results when the prior has little or no influence on the individual posterior distributions, such as in situations with a large number of observations per participant. Of course, if the individual parameters are estimated accurately and precisely, the attenuation of the correlation tends to be negligible.

Our literature review showed that nearly 50% of the articles published in the 2012 volume of the *Journal of Experimental Psychology: General* reported at least one Pearson product-moment correlation coefficient. Despite the wide-spread use of correlations, most researchers fail to acknowledge that the correlation computed between noisy observations or uncertain parameter estimates often underestimates the true strength of the relationship between the variables. Here we outlined a Bayesian method that allows researchers to infer the underlying correlation in the presence of measurement error and estimation uncertainty. Of course, the measurement error of the observations and the uncertainty of the parameter estimates are not always readily available. Also, our simulations confirmed that for relatively low underlying correlations, correcting the attenuation is likely to have only a small effect. We nevertheless encourage researchers to carefully consider the issue of attenuation and whenever sensible take into account the uncertainty that accompanies parameter estimation and psychological measurement.
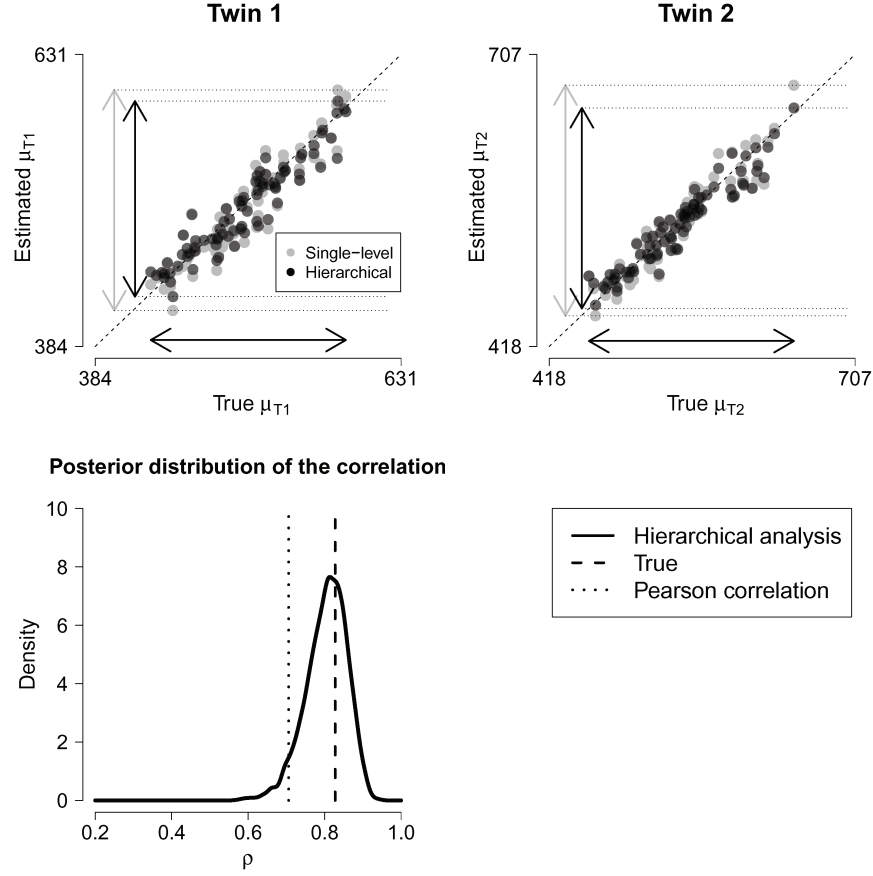
Figure 1. Bayesian hierarchical inference for the correlation in the presence of estimation uncertainty. The gray points show the point estimates $\hat{\mu}_{T1}$ and $\hat{\mu}_{T2}$ plotted against their true values. The black points show the posterior means of the inferred μ_{T1} and μ_{T2} parameters from the hierarchical analysis plotted against their true values. The horizontal black arrows indicate the range of the true values; the vertical gray and black arrows indicate the ranges of the point estimates and the inferred parameters, respectively. The inferred parameters are less dispersed than the point estimates. The bottom panel shows the posterior distribution of the correlation ρ from the hierarchical analysis: it more closely approximates the true correlation between μ_{T1} and μ_{T2} than the Pearson correlation coefficient r computed between the point estimates.
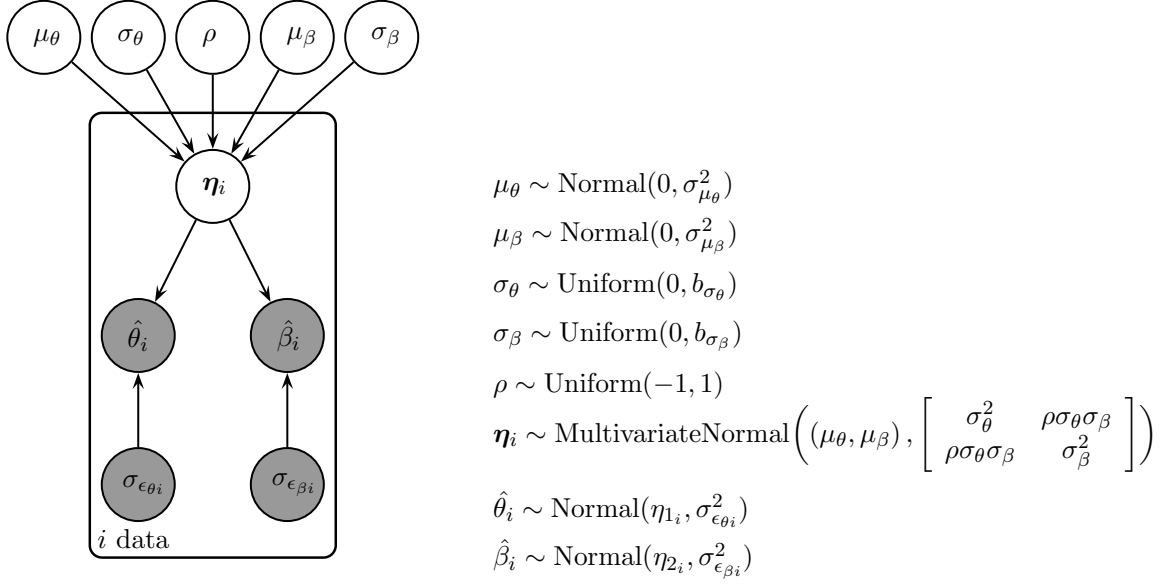
$$\mu_\theta \sim \text{Normal}(0, \sigma^2_{\mu_\theta})$$

$$\mu_\beta \sim \text{Normal}(0, \sigma^2_{\mu_\beta})$$

$$\sigma_\theta \sim \text{Uniform}(0, b_{\sigma_\theta})$$

$$\sigma_\beta \sim \text{Uniform}(0, b_{\sigma_\beta})$$

$$\rho \sim \text{Uniform}(-1, 1)$$

$$\boldsymbol{\eta}_i \sim \text{MultivariateNormal}\left((\mu_\theta, \mu_\beta), \begin{bmatrix} \sigma^2_\theta & \rho\sigma_\theta\sigma_\beta \\ \rho\sigma_\theta\sigma_\beta & \sigma^2_\beta \end{bmatrix}\right)$$

$$\hat{\theta}_i \sim \text{Normal}(\eta_{1_i}, \sigma^2_{\epsilon_{\theta i}})$$

$$\hat{\beta}_i \sim \text{Normal}(\eta_{2_i}, \sigma^2_{\epsilon_{\beta i}})$$

Figure 2. *Graphical model for inferring the correlation in the presence of measurement error.* Observed variables are represented by shaded nodes and unobserved variables are represented by unshaded nodes. The graph structure indicates dependencies between the nodes. The node $\boldsymbol{\eta} = (\theta, \beta)$ represents the unobserved true values, $\hat{\theta}$ and $\hat{\beta}$ represent the observed variables, and $\sigma^2_{\epsilon_\theta}$ and $\sigma^2_{\epsilon_\beta}$ represent the observed error variances. The nodes μ_θ and μ_β represent the unobserved group-level means, and σ_θ and σ_β represent the unobserved group-level standard deviations. The node ρ represents the unobserved true correlation between θ and β.
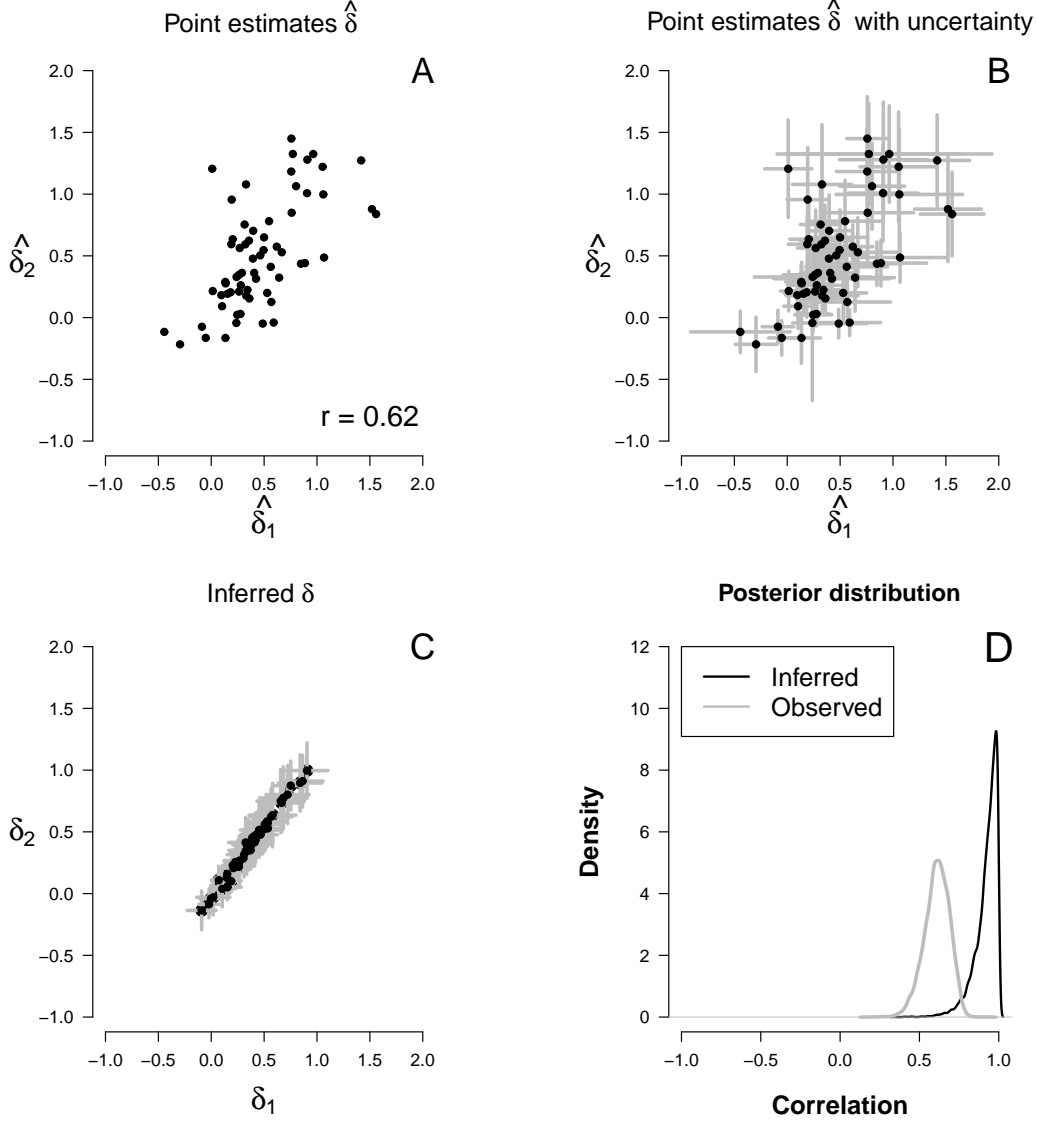
Figure 3.   Observed and inferred correlation between parameters of cumulative prospect theory (CPT). Panel A shows the Pearson correlation coefficient r and a scatterplot between the point estimates $\hat{\delta}_1$ and $\hat{\delta}_2$. Panel B shows the same scatterplot but includes the estimation uncertainties $\sigma_{\epsilon_{\delta 1}}$ and $\sigma_{\epsilon_{\delta 2}}$ (gray lines). Panel C shows a scatterplot between the posterior means of the inferred $\boldsymbol{\eta} = (\delta_1, \delta_2)$ parameters. The gray lines show the standard deviation of the posterior distributions. The posterior means of the inferred δ_1 and δ_2 parameters are shrunk toward their group mean and are associated more strongly than the point estimates. Panel D shows that the posterior distribution of the inferred correlation ρ (black density line) is shifted to higher values relative to the posterior of the observed correlation r (gray density line).
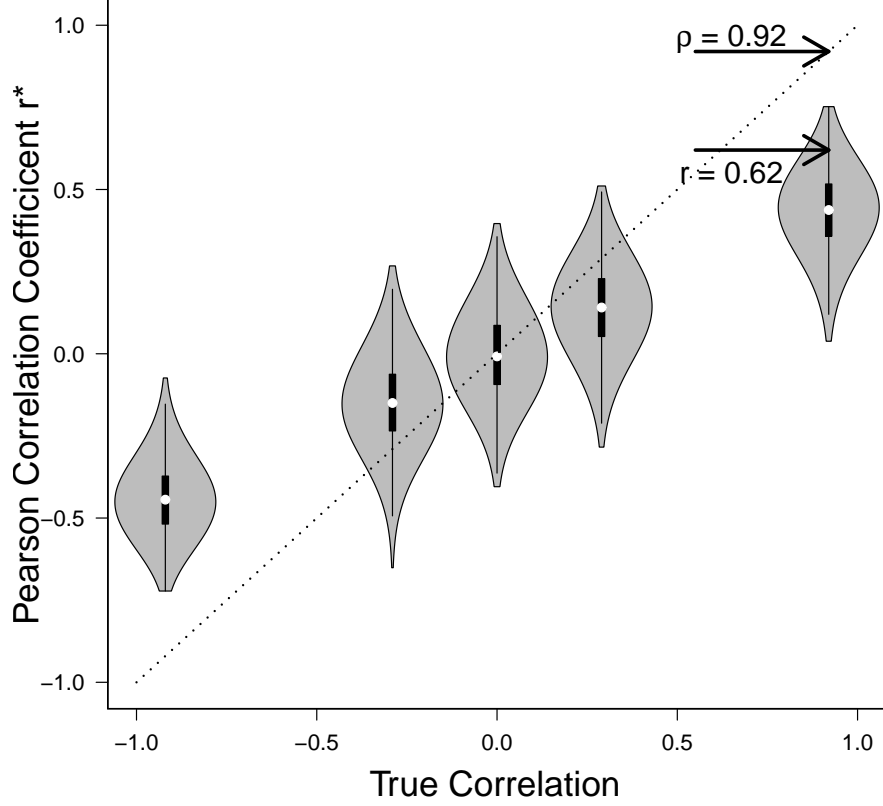
Figure 4. *Expected attenuation of the correlation between the δ parameters of cumulative prospect theory (CPT).* We conducted five sets of simulations, each with a different value of the true correlation. For each set, we generated $1,000$ synthetic data sets. The gray violin plots show the distribution of the $1,000$ Pearson correlation coefficients r^* for the five values of the true correlation. The r^* values were generated using the estimation uncertainties $\sigma^2_{\epsilon_{\delta 1}}$ and $\sigma^2_{\epsilon_{\delta 2}}$ and the posterior means of the group-level μ_{δ_1}, μ_{δ_2}, $\sigma^2_{\delta_1}$, and $\sigma^2_{\delta_2}$ parameters. The upper arrow shows the posterior mean of the inferred correlation ρ in the observed data. The lower arrow shows the Pearson correlation coefficient r between the point estimates $\hat{\delta}_1$ and $\hat{\delta}_2$ in the observed data. The figure shows that (1) the attenuation increases with the absolute value of the true correlation, and (2) the observed Pearson r between the point estimates is well within the 2.5^{th} and 97.5^{th} percentile of the r^* values predicted by the simulations.
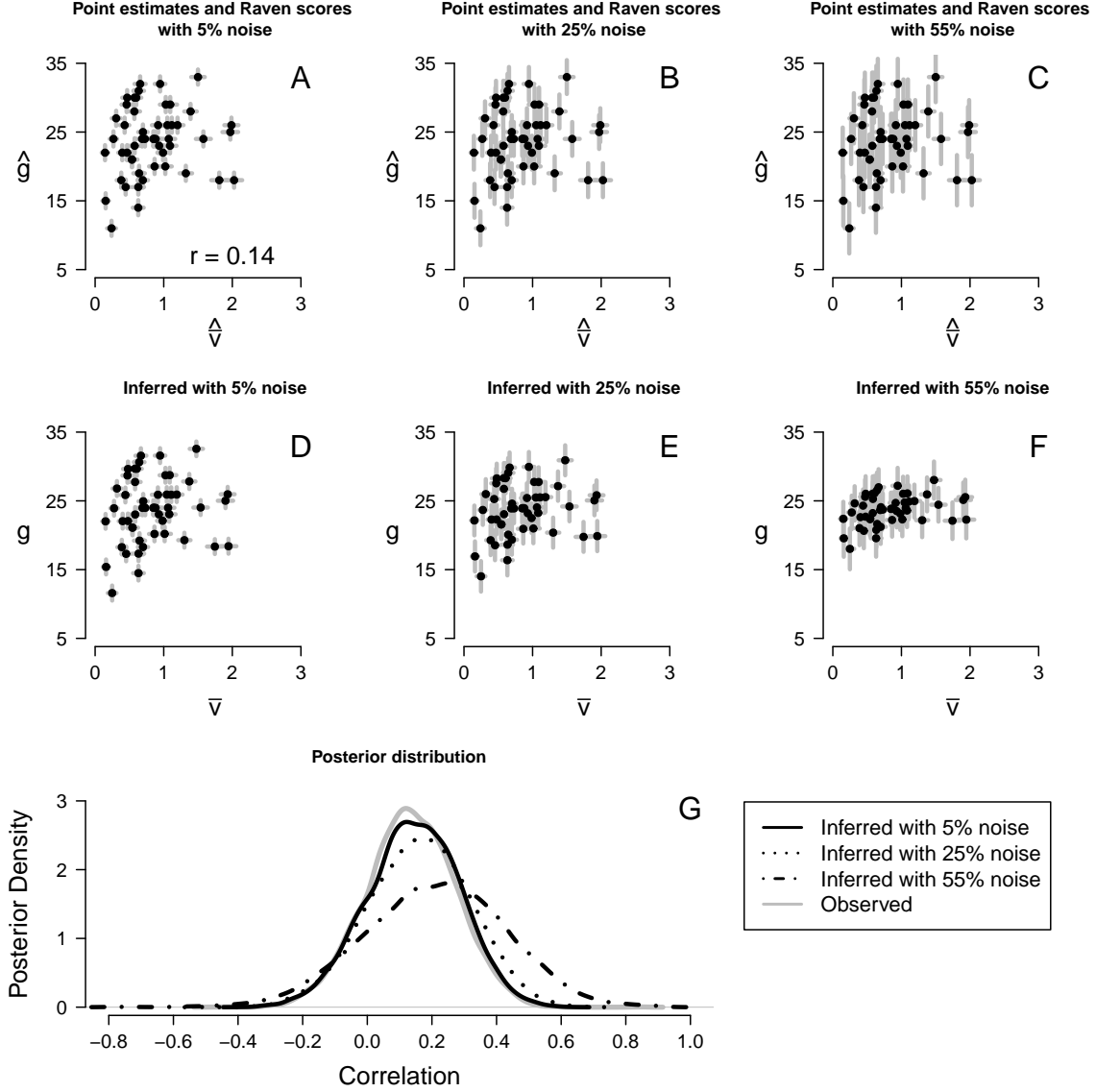
Figure 5. *Observed and inferred correlation between mean drift rate $\bar{v}$ and Raven total score g.* Panels A - C show scatterplots between the point estimates for mean drift rate $\hat{\bar{v}}$ and the observed Raven total score $\hat{g}$. The gray lines show the corresponding estimation uncertainties $\sigma_{\epsilon_{\bar{v}}}$ and the standard deviation of the measurement error σ_{ϵ_g} for the 5%, 25%, and 55% measurement noise scenarios. Panel A also shows the Pearson correlation coefficient r. Panels D - F show scatterplots between the posterior means of the inferred $\eta = (\bar{v}, g)$ parameters for the three scenarios. The gray lines show the standard deviations of the posterior distributions. The posterior means of the inferred $\bar{v}$ and g parameters are shrunk toward their group mean, where the degree of shrinkage is determined by the amount of measurement noise assumed for g. Panel G shows the posterior distribution of the observed (r; gray density line) and the inferred correlations (ρ) for the 5% (solid black density line), 25% (dotted black density line,), and 55% (dotted-dashed black density line) measurement noise scenarios. Note the only negligible shift is the mean of the posterior distribution of the inferred correlation ρ as a function of the amount of measurement noise.
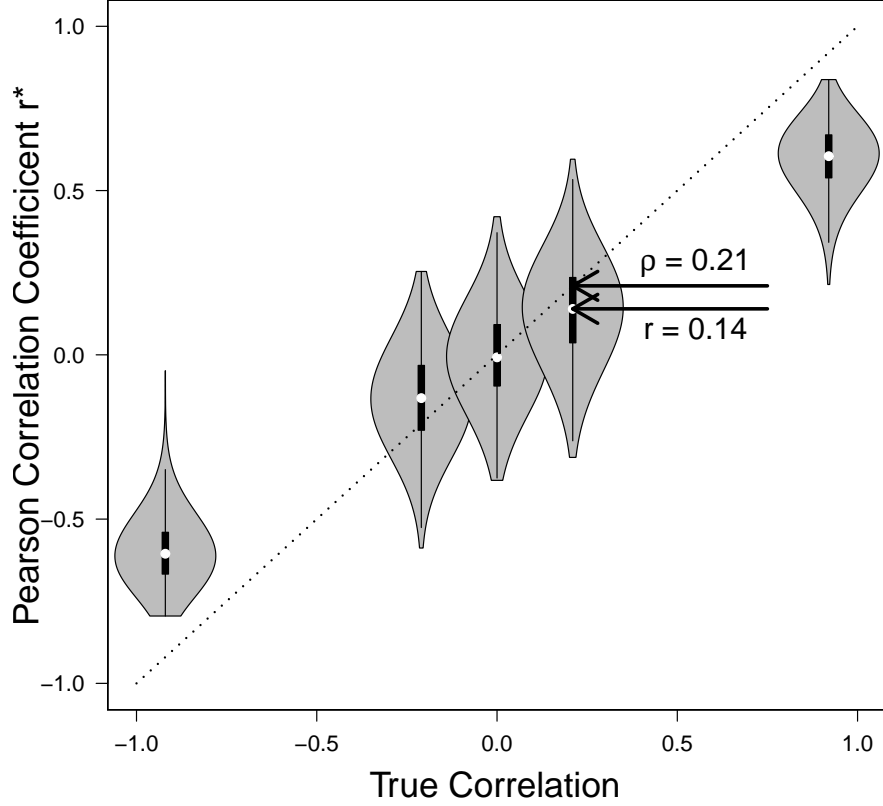
Figure 6.   Expected attenuation of the correlation between mean drift rate $\bar{v}$ and Raven total score g. We conducted five sets of simulations, each with a different value of the true correlation. For each set, we generated $1,000$ synthetic data sets. The gray violin plots show the distribution of the $1,000$ Pearson correlation coefficients r^* for the five values of the true correlation. The r^* values were generated assuming 55% error variance in the Raven scores, using the estimation uncertainties $\sigma^2_{\epsilon_{\bar{v}}}$ and the posterior means of the group-level $\mu_{\bar{v}}$, μ_g, $\sigma^2_{\bar{v}}$, and σ^2_g parameters. The upper arrow shows the posterior mean of the inferred correlation ρ in the observed data. The lower arrow shows the Pearson correlation coefficient r between the point estimates $\hat{\bar{v}}$ and $\hat{g}$ in the observed data. The figure shows that the median (white circle) of the Pearson correlation coefficients r^* predicted by the simulations very closely approximates the observed Pearson r between the point estimates.

References

Behseta, S., Berdyyeva, T., Olson, C. R., & Kass, R. E. (2009). Bayesian correction for attenuation of correlation in multi-trial spike count data. *Journal of Neurophysiology*, *101*, 2186–2193.

Buonaccorsi, J. P. (2010). *Measurement error: Models, methods, and applications.* Boca Raton, FL: Chapman & Hall/CRC.

Charles, E. P. (2005). The correction for attenuation due to measurement error: Clarifying concepts and creating confidence sets. *Psychological Methods*, *10*, 206-226.

Cheng, C.-L., & Van Ness, J. W. (1999). *Statistical regression with measurement error.* London, UK: Arnold.

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, *19*, 300-315.

Congdon, P. (2006). *Bayesian statistical modelling* (2nd ed.). Chichester, UK: John Wiley & Sons Ltd.

Dickey, J. M., & Lientz, B. P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *The Annals of Mathematical Statistics*, *41*, 214–226.

Farrell, S., & Ludwig, C. J. H. (2008). Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychonomic Bulletin & Review*, *15*, 1209–1217.

Fuller, W. A. (1987). *Measurement error models.* New York, NY: John Wiley & Sons, Inc.

Gamerman, D., & Lopes, H. F. (2006). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference.* Boca Raton, FL: Chapman & Hall/CRC.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* Cambridge, UK: Cambridge University Press.

Gelman, A., Meng, X., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*, 733–807.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice.* Boca Raton, FL: Chapman & Hall/CRC.

Glöckner, A., & Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition*, *123*, 21–32.

Gustafson, P. (2004). *Measurement error and misclassification in statistics and epidemiology: Impacts and Bayesian adjustments.* Boca Raton, FL: Chapman & Hall/CRC.

Hamel, R., & Schmittmann, V. D. (2006). The 20-minute version as a predictor of the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement*, *66*, 1039–1046.

Hintze, J. L., & Nelson, R. D. (1998). Violin plots: A box plot-density trace synergism. *American Statistician*, *52*, 181–184.

Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, *69*, 382-386.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* New York, NY: Praeger.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*, 70–98.

Kruschke, J. K. (2010). *Doing Bayesian data analysis: A tutorial introduction with R and BUGS.* Burlington, MA: Academic Press.

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, *55*, 1–7.

Lee, M. D., & Vanpaemel, W. (2016). Determining informative priors for cognitive models. *Manuscript submitted for publication*.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian modeling for cognitive science: A practical course.* Cambridge, UK: Cambridge University Press.

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach.* New York, NY: John Wiley & Sons.

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis.* Boca Raton, FL: Chapman & Hall/CRC.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, *28*, 3049–3067.

Ly, A., Marsman, M., & Wagenmakers, E.-J. (2015). Analytic posteriors for Pearson's correlation coefficient. *arXiv preprint arXiv:1510.01188*.

Ly, A., Verhagen, A., & Wagenmakers, E.-J. (2016a). An evaluation of alternative methods for testing hypotheses, from the perspective of Harold Jeffreys. *Journal of Mathematical Psychology*, *72*, 43–55. doi: http://dx.doi.org/10.1016/j.jmp.2016.01.003

Ly, A., Verhagen, A., & Wagenmakers, E.-J. (2016b). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, *72*, 19–32. doi: http://dx.doi.org/10.1016/j.jmp.2015.06.004

Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E.-J. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, *80*, 205-235.

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, *16*, 798–817.

Morey, R. D., & Wagenmakers, E.-J. (2014). Simple relation between Bayesian order-restricted and point-null hypothesis tests. *Statistics and Probability Letters*, *92*, 121-124.

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement*, *56*, 63–75.

Muff, S., Riebler, A., Held, L., Rue, H., & Saner, P. (in press). Bayesian analysis of measurement error models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Applied Statistics, Series C*.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*, 90–100.

Nilsson, H., Rieskamp, J., & Wagenmakers, E.-J. (2011). Hierarchical Bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, *55*, 84–93.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In K. Hornik, F. Leisch, & A. Zeileis (Eds.), *Proceedings of the 3rd international workshop on distributed statistical computing.* Vienna, Austria.

Plummer, M. (2013). JAGS Version 3.1.0 [Computer software manual]. Retrieved from `http://mcmc-jags.sourceforge.net/`

R Core Team. (2012). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/` (ISBN 3-900051-07-0)

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.

Ratcliff, R., Schmiedek, F., & McKoon, G. (2008). A diffusion model explanation of the worst performance rule for reaction time and IQ. *Intelligence*, *36*, 10–17.

Ratcliff, R., & Strayer, D. (2014). Modeling simple driving tasks with a one-boundary diffusion model. *Psychonomic Bulletin & Review*, *21*, 577–589.

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, *60*, 127–157.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 4: The Advanced Progressive Matrices*. San Antonio, TX: Harcourt Assessment.

Rouder, J. N., Lu, J., Morey, R. D., Sun, D., & Speckman, P. L. (2008). A hierarchical process-dissociation model. *Journal of Experimental Psychology: General*, *137*, 370–389.

Rouder, J. N., Lu, J., Speckman, P., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review*, *12*, 195–223.

Rouder, J. N., Lu, J., Sun, D., Speckman, P., Morey, R., & Naveh-Benjamin, M. (2007). Signal detection models with random participant and item effects. *Psychometrika*, *72*, 621–642.

Rouder, J. N., Sun, D., Speckman, P. L., Lu, J., & Zhou, D. (2003). A hierarchical Bayesian statistical framework for response time distributions. *Psychometrika*, *68*, 589–606.

Schmiedek, F., Oberauer, K., Wilhelm, O., Suss, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, *136*, 414–429.

Song, X.-Y., & Lee, S.-Y. (2012). A tutorial on the Bayesian approach for analyzing structural equation models. *Journal of Mathematical Psychology*, *56*, 135–148.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*, 72–101.

Stan Development Team. (2012). Stan modeling language [Computer software manual]. Retrieved from `http://mc-stan.org/`

Sturtz, S., Ligges, U., & Gelman, A. E. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software*, *12*, 1–16.

Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging*, *18*, 415–429.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.

Vanpaemel, W., & Lee, M. (2012). Using priors to formalize theory: Optimal attention and the generalized context model. *Psychonomic Bulletin & Review*, *19*, 1047-1056.

van Ravenzwaaij, D., Brown, S., & Wagenmakers, E.-J. (2011). An integrated perspective on the relation between response speed and intelligence. *Cognition*, *119*, 381–393.

Vickers, D., Nettelbeck, T., & Willson, R. J. (1972). Perceptual indices of performance: The measurement of "inspection time" and "noise" in the visual system. *Perception*, *1*, 263–295.

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, *32*, 1206–1220.

Wabersich, D., & Vandekerckhove, J. (2014). Extending JAGS: A tutorial on adding custom distributions to JAGS (with a diffusion model example). *Behavior Research Methods*, *46*, 15–28.

Wagenmakers, E.-J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology*, *21*, 641–671.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*, 158–189.

Wetzels, R., Grasman, R. P. P. P., & Wagenmakers, E.-J. (2010). An encompassing prior generalization of the Savage-Dickey density ratio test. *Computational Statistics & Data Analysis*, *54*, 2094–2102.

Appendix
WinBUGS Implementation

```
{
  # Data
  for (i in 1:N){
    # eta[i,1] = theta[i]; eta[i,2] = beta[i];
    # mu[1] = mu_theta; mu[2] = mu_beta;
    # ISigma_cov = inverse of Sigma_cov matrix
    eta[i,1:2] ~ dmnorm(mu[],ISigma_cov[,])

    # observed[i,1] = theta_hat[i]; observed[i,2] = beta_hat[i];
    # Isigma_epsilon[i,1] = Inverse of sigma_epsilon^2_theta[i];
    # Isigma_epsilon[i,2] = Inverse of sigma_epsilon^2_beta[i]
    for (j in 1:2){
      observed[i,j] ~ dnorm(eta[i,j],Isigma_epsilon[i,j])
    }
  }

  # Priors
  mu[1] ~ dnorm(0,prec_mu_theta)
  mu[2] ~ dnorm(0,prec_mu_beta)

  sigma_theta ~ dunif(0,b_sigma_theta)
  sigma_beta ~ dunif(0,b_sigma_beta)

  rho ~ dunif(-1,1)

  # Reparameterization
  Sigma_cov[1,1] <- pow(sigma_theta,2)
  Sigma_cov[1,2] <- rho*sigma_theta*sigma_beta
  Sigma_cov[2,1] <- rho*sigma_theta*sigma_beta
  Sigma_cov[2,2] <- pow(sigma_beta,2)
  ISigma_cov[1:2,1:2] <- inverse(Sigma_cov[1:2,1:2])
}
```

The R code that calls the WinBUGS code using the R2WinBUGS (Sturtz, Ligges, & Gelman, 2005) package is available in the supplemental materials at `http://dora.erbe -matzke.com/publications.html`. The R code allows users to adjust the precision (i.e., inverse variance) of the normal prior distribution of `mu[1]` and `mu[2]` by specifying the value of `prec_mu_theta` and `prec_mu_beta`. Similarly, the R code allows users to adjust the range of the uniform prior distribution of `sigma_theta` and `sigma_beta` by specifying the value of `b_sigma_theta` and `b_sigma_beta`.